DDi
Drug Development informatics

WHITE PAPER

Handling Missing Data in Clinical trials

Ms. Ankita Mistry, Dr. A. K. Mathai

www.makrocare.com/DDi

In the recent past the number of clinical trials have been increased dramatically world over. There are a lot of regulations in data analysis and reporting the results. However, sometimes, statisticians pose difficulty in handling missing observations generated from clinical trials. Missing observations may happen due to various factors such as the subject could not come for follow up in a particular time point or it could be due to migration of patients from the place of treatment or it could be due to treatment failure and subject would have gone to treatment elsewhere. Thus the reasons could be anything which is beyond the control of investigator or sponsor or CRO. However, it is the responsibility of the statistician to perform the analysis in a proper way to take into account of the missing observations while doing the final analysis. The objective of this paper is to highlight the importance of handling of the missing data and a brief description of the procedures for the same.

Impact of Missing Data in Clinical Trial

Missing data are a major problem in clinical trial especially, when data collection is costly. This is more critical when the disease condition is very rare and number of eligible subjects is very less for final analysis. In case of interpreting the Clinical trial results, it is always creates a problem when the number of missing values are substantial. Ultimately it will affect the statistical power of the test and the results may not be convincing to the scientific community. Moreover, the missing data is a potential source of bias. In this situation the statistician has to adopt a proper technique to handle missing data. He should come out with a result which should be convincing and reliable to the scientific community.

Reasons and Degree of Missing Data

There are many possible reasons for missing data such as patient refusal to continue in the study, treatment failures, early successes of the treatment, adverse events, patient migration, withdrawal from the trial due to various reasons, drop-out from the trial by the participant, not recording the data at the source due to personal lapse or instrumental failure etc., death of the study participant due to accident or any other reason. Thus the reason for missing data varies widely. The degree of missing data can occur like missing at baseline or at one or several follow-up assessments and even if a patient completes the study, some data may remain uncollected.

Classification of Missing Data

Usually there are three classification of missing data. If the probability of an observation being missing does not depend on observed or unobserved measurements then the missing observation is classified as Missing Completely At Random (MCAR). For example, a patient has been migrated to another city for non-health reasons, then, this patient would be considered as drop-out of a study. This patient's data may be considered as missing completely at random. However, the demographic and baseline characteristics suppose to be similar for both study completed cases and drop-out cases. The second scenario is that if the probability of an observation is missing depends only on observed data and not depends on unobserved data, and then the observation is Missing At Random (MAR). For example, when a patient drops out due to lack of efficacy reflected by a series of poor efficacy outcomes that have been observed, the appropriate value to assign to the subsequent efficacy endpoint for this patient can be calculated using the observed data. The third scenario is that when the observations are missing neither following MCAR pattern nor MAR pattern, then it is classified as Missing Not At Random (MNAR). It is also called non-ignorable. For example, a participant in a weight-loss study does not attend a weigh-in due to concerns about his/her weight loss, his/her data are missing due to non-ignorable factors.

Procedures for Handling Missing Data

The most common approach of handling missing data is imputation methods. The purpose of imputation is to produce a complete data set which can then be analyzed using standard statistical methods. The observed values are used to impute values for the missing observations. There are two kinds of imputation methods available, viz., the Single Imputation and Multiple Imputation methods. However, the multiple imputation method is slightly complex and the

discussion about the same will be beyond the scope of this paper. The Single imputation method consists of foursub-categories as follows:

- **Mean Imputation:** The sample mean of a variable replaces any missing data for that variable
- **Hot-deck Imputation:** Missing values are replaced with values taken from matching respondents
- **Last observation Carried Forward (or LOCF):** The last observed value is used to fill in missing values at subsequent points in a longitudinal study. This method is very often used in clinical research
- **Predicted Mean:** Ordinary Least-Squares multiple regression algorithm is used to impute the most likely value. In this method, researchers develop a regression equation based on complete case data for a given variable, treating it as the outcome and using all other relevant variables as predictors

Conclusion and Recommendation

Missing observations in clinical trial is unavoidable. However, it needs to be minimized as far as possible. The Sponsor, investigators and CRO must take precautionary measures to reduce the missing data in clinical trial. But once the missing data has been obtained from the clinical trial, the statistician must employ the appropriate procedures to handle the same. He should not simply ignore the missing data from the final analysis. However, the handling of missing data procedures must be clearly mentioned in the protocol. Thus the statistician can come out with credible results which would be acceptable for the regulatory authorities. It will facilitate to obtain the marketing approval from the concerned agencies.

About DDi

DDi (Drug Development informatics), is an IT division of MakroCare that provides Clinical Technologies and IT solutions to the Life Sciences industry. With a unique blend of functional, domain and technology expertise, DDi is a prominent partner to provide techno-functional products & services.